RESEARCH ARTICLE

Breast Cancer Association Studies in a Han Chinese Population using 10 European-ancestry-associated Breast Cancer Susceptibility SNPs

Yan-Ping Guan[2&], Xue-Xi Yang[1&], Guang-Yu Yao[2], Fei Qiu[1], Jun Chen[2], Lu-jia Chen[2], Chang-Sheng Ye[2*], Ming Li[1*]

Abstract

Background: Genome-wide association studies (GWAS) have identified various genetic susceptibility loci for breast cancer based mainly on European-ancestry populations. Differing linkage disequilibrium patterns exist between European and Asian populations. Methods: Ten SNPs (rs2075555 in COL1A1, rs12652447 in FBXL17, rs10941679 in 5p12/MRPS30, rs11878583 in ZNF577, rs7166081 in SMAD3, rs16917302 in ZNF365, rs311499 in 20q13.3, rs1045485 in CASP8, rs12964873 in CDH1 and rs8170 in 19p13.1) were here genotyped in 1009 Chinese females (487 patients with breast cancer and 522 control subjects) using the Sequenom MassARRAY iPLEX platform. Association analysis based on unconditional logistic regression was carried out to determine the odds ratio (OR) and 95% confidence interval (95% CI) for each SNP. Stratification analyses were carried out based on the estrogen receptor (ER) and progesterone receptor (PR) status. Results: Among the 10 SNPs, rs10941679 showed significant association with breast cancer when differences between the case and control groups in this Han Chinese population were compared (30.09% GG, 45.4% GA and 23.7% AA; *P* = 0.012). Four SNPs (rs311499, rs1045485, rs12964873 and rs8170) showed no polymorphisms in our study. The remaining five SNPs showed no association with breast cancer in the present population. Immunohistochemical tests showed that rs2075555 was associated with ER status; the AA genotype showed greater association with ER negative than ER positive (OR = 0.54, 95% CI, 0.29–0.99; *P* = 0.046). AA of rs7166081 was also associated with ER status, but showed a greater association with ER positive than negative (OR = 1.59, 95% CI = 1.04–2.44; *P* = 0.031). However, no significant associations were found among the SNPs and PR status. Conclusion: In this study using a Han Chinese population, rs10941679 was the only SNP associated with breast cancer risk, indicating a difference between European and Chinese populations in susceptibility loci. Therefore, confirmation studies are necessary before utilization of these loci in Chinese.

Keywords: Breast cancer - single nucleotide polymorphisms - European susceptibility - Chinese population

Asian Pac J Cancer Prev, **15 (1)**, 85-91

Introduction

Breast cancer is one of the most common malignancies in females, and causes serious health damage to females worldwide. In European and American countries, the incidence of breast cancer is ~14% (Jemal et al., 2006). In China, the incidence of breast cancer is increasing annually (Yu et al., 2007). Moreover, compared to European and American patients, the onset of breast cancer begins 10 to 15 years earlier in Chinese patients.

The cause of breast cancer has not yet been fully elucidated. Embryonic mutations account for about 20–40% of the hereditary breast cancer cases (Walsh et al., 2007). Sporadic breast cancer is often caused by genetic susceptibility and environmental factors that interact with each other. Because various mutations explain a proportion of breast cancers, widespread genetic polymorphisms in a population results in individuals with differing breast cancer susceptibility risks. Gene polymorphisms result in various responses when humans encounter cancer-causing toxins, which can inhibit or accelerate the occurrence of breast cancer. Genetic polymorphisms are a key component of the differences in breast cancer susceptibilities among individuals. Therefore, such loci polymorphisms could be used as biomarkers for developing individualized prevention strategies. While the role of one susceptibility locus might be limited, multiple loci might have a cascading effect; for example, the risk of breast cancer is increased by approximately six times in females carrying 14 common risk alleles compared to those carrying one (Pharoah et al., 2008). Therefore, susceptibility gene polymorphism loci play an important role in breast cancer. Genome-wide association studies (GWAS) have identified various genetic susceptibility

[1]School of Biotechnology, [2]Breast Center Nanfang Hospital, Southern Medical University, Guangzhou, China [&]Equal contributors
**For correspondence: mingli2006_2006@126.com, yechsh2006@126.com*

Table 1. Estimation of the Statistical Power of the Unmatched Case-control Study

Subjects	MAF	Statistical power									
		OR=1.1	OR=1.2	OR=1.3	OR=1.4	OR=1.5	OR=1.6	OR=1.7	OR=1.8	OR=1.9	OR=2.0
	0.03	0.0661	0.1127	0.1876	0.2862	0.4004	0.5194	0.6328	0.7324	0.8138	0.8761
487 cases	0.06	0.0814	0.173	0.3163	0.488	0.6552	0.7915	0.8865	0.9441	0.9749	0.9897
	0.09	0.0957	0.229	0.4275	0.6375	0.8057	0.9116	0.9656	0.9884	0.9966	0.9991
522 controls	0.12	0.1091	0.2801	0.5195	0.7421	0.8891	0.9614	0.989	0.9974	0.9995	0.9999
	0.15	0.1215	0.3258	0.594	0.8136	0.9347	0.9822	0.9961	0.9993	0.9999	0.9999
	0.18	0.1328	0.3661	0.6534	0.8622	0.9599	0.9912	0.9985	0.9998	0.9999	0.9999
	0.21	0.1431	0.4013	0.7003	0.8954	0.9741	0.9953	0.9994	0.9999	0.9999	0.9999
	0.24	0.1523	0.4316	0.7373	0.9182	0.9824	0.9973	0.9997	0.9999	0.9999	0.9999
	0.27	0.1604	0.4574	0.7661	0.934	0.9874	0.9983	0.9998	0.9999	0.9999	0.9999
	0.3	0.1675	0.4789	0.7884	0.9451	0.9905	0.9988	0.9999	0.9999	0.9999	0.9999

MAF, Minor Allele Frequency; OR, Odds Ratio

Table 2. The Detail Information of Breast Cancer Case-control Group Including All Subjects

Variable	Cases	Controls
All subjects	487	522
Age, years (mean ± SD)		
Guangdong-Shandong	48.20±9.850 (range 22-80)	46.99 ± 9.647 (range 20-78)
Guangdong	47.98±11.022 (range 22-80)	47.15±9.820 (range 25-76)
Shandong	48.42±8.482 (range 30-73)	46.44±7.732 (range 35-60)
Receptor status		
Estrogen receptor		
Guangdong-Shandong (n=487)		
Positive	288(59.1%)	
Negative	199(40.9%)	
Guangdong (n=248)		
Positive	141(56.9%)	
Negative	107(43.1%)	
Shandong (n=239)		
Positive	147(61.5%)	
Negative	92(38.5%)	
Progesterone receptor		
Guangdong-Shandong (n = 487)		
Positive	273(56.1%)	
Negative	214(43.9%)	
Guangdong (n=248)		
Positive	128(51.6%)	
Negative	120(48.4%)	
Shandong (n=239)		
Positive	145(60.7%)	
Negative	94(39.3%)	

loci for breast cancer; however, most study populations have been of European ancestry (Raskin et al., 2008; Stacey et al., 2008; Walker et al., 2010; Fergus et al., 2012). Differing linkage disequilibrium patterns and environmental factors exist between European and Asian populations, and thus GWAS-identified single nucleotide polymorphisms (SNPs) in one population may not show significance in another (McCracken et al., 2007; Long et al., 2010). Thus, a suitable breast cancer assessment risk model must be developed in Chinese female patients.

This study evaluated associations between variant loci and breast cancer using rs2075555 in COL1A1, rs12652447 in FBXL17, rs10941679 in 5p12/MRPS30, rs11878583 in ZNF577, rs7166081 in SMAD3, rs16917302 in ZNF365, rs311499 in 20q13.3, rs1045485 in CASP8, rs12964873 in CDH1 and rs8170 in 19p13.1, which have been previously reported in GWAS of European-ancestry populations. GWAS must be conducted in non-European populations to determine the genetic basis of breast cancer susceptibility.

Materials and Methods

Study subjects

A total of 487 breast cancer cases and 522 cancer-free controls were recruited from Nanfang Hospital of Southern Medical University, Guangdong, China (248 female patients) and Affiliated Hospital of Medical College of Qingdao University, Shandong, China (239 female patients) from 2009 to 2012. All breast cancer cases were confirmed histopathologically without restriction of age or histological type, and no patients had a history of other cancers. The histological information included estrogen receptor (ER) and progesterone receptor (PR) status, as determined immunohistochemically. Cancer-free controls were recruited randomly from Nanfang Hospital (306 females) and Affiliated Hospital of Qingdao University (216 females). All subjects were unrelated Chinese females of Han ancestry.

Genotyping

After obtaining informed consent, 5-ml peripheral blood samples were collected and delivered to the laboratory in the frozen state. Genomic DNA was extracted from a 200-μl peripheral blood sample using the Tiangen™ Genomic DNA Kit (Tiangen, Beijing, China) according to the manufacturer's instructions and stored at –80°C until use. Ten SNPs were selected based on data from the 1000 Genomes Project (http://www.1000genomes.org/). All of the SNPs were genotyped using the SEQUENOM MassARRAY matrix-assisted laser desorption ionization–time of flight mass spectrometry platform (Sequenom, San Diego, California, USA). Primers were designed using a semiautomated method (Assay Design 3.1, Sequenom). The call rate for each assay was set at >90%.

Statistical analysis

The Hardy–Weinberg equilibrium (HWE) was calculated using Haploview 4.2 (Daly Lab, Cambridge, MA). The genotype and allele distributions between case and control subjects were compared. Association analysis based on unconditional logistic regression after

Table 3. Distributions of 6 SNPs in Breast Cancer and Control Groups

Gene/Chr	SNP	Model	Genotype	Control	Case	OR (95% CI)	*P*-value
COL1A1	rs20775555	Overdominant	C/C-A/A	297 (57.2%)	263 (54.3%)	1	0.36
			C/A	222 (42.8%)	221 (45.7%)	1.12 (0.88-1.44)	
		Recessive	C/C-C/A	452 (87.1%)	438 (90.5%)	1	0.087
			A/A	67 (12.9%)	46 (9.5%)	0.71 (0.48-1.05)	
		Dominant	C/C	230 (44.3%)	217 (44.8%)	1	0.87
			C/A-A/A	289 (55.7%)	267 (55.2%)	0.98 (0.76-1.26)	
		Codominant	C/C	230 (44.3%)	217 (44.8%)	1	0.21
			C/A	222 (42.8%)	221 (45.7%)	1.06 (0.81-1.37)	
			A/A	67 (12.9%)	46 (9.5%)	0.73 (0.48-1.11)	
FBXL17	rs12652447	Overdominant	A/A-G/G	282 (54.2%)	253 (52.2%)	1	0.51
			A/G	238 (45.8%)	232 (47.8%)	1.09 (0.85-1.39)	
		Recessive	A/A-A/G	417 (80.2%)	383 (79%)	1	0.63
			G/G	103 (19.8%)	102 (21%)	1.08 (0.79-1.47)	
		Dominant	A/A	179 (34.4%)	151 (31.1%)	1	0.27
			A/G-G/G	341 (65.6%)	334 (68.9%)	1.16 (0.89-1.51)	
		Codominant	A/A	179 (34.4%)	151 (31.1%)	1	0.54
			A/G	238 (45.8%)	232 (47.8%)	1.16 (0.87-1.53)	
			G/G	103 (19.8%)	102 (21%)	1.17 (0.83-1.66)	
MRPS30	rs10941679	Overdominant	G/G-A/A	175 (45.3%)	214 (54.6%)	1	0.0098
			G/A	211 (54.7%)	178 (45.4%)	0.69 (0.52-0.91)	
		Recessive	G/G-G/A	297 (76.9%)	299 (76.3%)	1	0.83
			A/A	89 (23.1%)	93 (23.7%)	1.04 (0.74-1.45)	
		Dominant	G/G	86 (22.3%)	121 (30.9%)	1	0.0066
			G/A-A/A	300 (77.7%)	271 (69.1%)	0.64 (0.47-0.89)	
		Codominant	G/G	86 (22.3%)	121 (30.9%)	1	0.012
			G/A	211 (54.7%)	178 (45.4%)	0.60 (0.43-0.84)	
			A/A	89 (23.1%)	93 (23.7%)	0.74 (0.50-1.11)	
ZNF577	rs11878583	Overdominant	A/A-G/G	309 (59.2%)	282 (57.9%)	1	0.68
			A/G	213(40.8%)	205 (42.1%)	1.05 (0.82-1.36)	
		Recessive	A/A-A/G	467 (89.5%)	440 (90.3%)	1	0.64
			G/G	55 (10.5%)	47 (9.7%)	0.91 (0.60-1.37)	
		Dominant	A/A	254 (48.7%)	235 (48.2%)	1	0.9
			A/G-G/G	268 (51.3%)	252 (51.8%)	1.02 (0.79-1.30)	
		Codominant	A/A	254 (48.7%)	235 (48.2%)	1	0.86
			A/G	213 (40.8%)	205 (42.1%)	1.04 (0.80-1.35)	
			G/G	55 (10.5%)	47 (9.7%)	0.92 (0.60-1.42)	
SMAD3	rs7166081	Overdominant	G/G-A/A	297 (58.5%)	260 (53.8%)	1	0.14
			A/G	211 (41.5%)	223 (46.2%)	1.21 (0.94-1.55)	
		Recessive	G/G-A/G	380 (74.8%)	357 (73.9%)	1	0.75
			A/A	128 (25.2%)	126 (26.1%)	1.05 (0.79-1.39)	
		Dominant	G/G	169 (33.3%)	134 (27.7%)	1	0.059
			A/G-A/A	339 (66.7%)	349 (72.3%)	1.30 (0.99-1.70)	
		Codominant	G/G	169 (33.3%)	134 (27.7%)	1	0.15
			A/G	211 (41.5%)	223 (46.2%)	1.33 (0.99-1.79)	
			A/A	128 (25.2%)	126 (26.1%)	1.24 (0.89-1.73)	
NF365	rs16917302	Overdominant	A/A-C/C	340 (65.4%)	321 (66.5%)	1	0.72
			C/A	180 (34.6%)	162 (33.5%)	0.95 (0.73-1.24)	
		Recessive	A/A-C/A	499 (96%)	460 (95.2%)	1	0.58
			C/C	21 (4%)	23 (4.8%)	1.19 (0.65-2.18)	
		Dominant	A/A	319 (61.4%)	298 (61.7%)	1	0.91
			C/A-C/C	201 (38.6%)	185 (38.3%)	0.99 (0.76-1.27)	
		Codominant	A/A	319 (61.4%)	298 (61.7%)	1	0.82
			C/A	180 (34.6%)	162 (33.5%)	0.96 (0.74-1.26)	
			C/C	21 (4%)	23 (4.8%)	1.17 (0.64-2.16)	

The *P* value is counted by the web-based tool SNPstats, the corresponding OR is counted after age adjustment

age-adjustment in case-control study was carried out by calculating the odds ratio (OR) and 95% confidence interval (CI) for each SNP in codominant and dominant genetic models.

Further stratifying analysis was conducted based on data ER or PR status. The statistical tests were implemented using the web-based tool SNPstats (http://bioinfo.iconcologia.net/SNPstats), and the significance level was set at 0.05.

Results

Quality control for all SNPs with Hardy-Weinberg Equilibrium (HWE) $P>0.05$ in healthy controls,

Table 4. SNPs that were Associated with ER-positive or ER-negative Tumours

				ER Positive			ER Negative			
SNP	Model	Genotype	Controls, N(%)	Cases, N (%)	OR(95% CI)	*P*	Cases, N (%)	OR(95% CI)	*P*	*P*-value
rs20775555	Overdominant	C/C-A/A	297 (57.6%)	154 (53.5%)	1		109 (55.6%)	1		0.046
		C/A	219 (42.4%)	134 (46.5%)	1.18 (0.88-1.58)	0.26	87 (44.4%)	1.08 (0.78-1.51)	0.64	
	Recessive	C/C-C/A	449 (87%)	267 (92.7%)	1		171 (87.2%)	1		0.64
		A/A	67 (13%)	21 (7.3%)	0.53 (0.32-0.88)	0.011	25 (12.8%)	0.98 (0.60-1.60)	0.94	
	Dominant	C/C	230 (44.6%)	133 (46.2%)	1		84 (42.9%)	1		0.47
		C/A-A/A	286 (55.4%)	155 (53.8%)	0.94 (0.70-1.25)	0.66	112 (57.1%)	1.07 (0.77-1.49)	0.68	
	Codominant	C/C	230 (44.6%)	133 (46.2%)	1		84 (42.9%)	1		0.14
		C/A	219 (42.4%)	134 (46.5%)	1.06 (0.78-1.43)		87 (44.4%)	1.09 (0.76-1.55)		
		A/A	67 (13%)	21 (7.3%)	0.54 (0.32-0.93)	0.036	25 (12.8%)	1.02 (0.61-1.72)	0.89	
rs12652447	Overdominant	A/A-G/G	281 (54.4%)	150 (52.1%)	1		103 (52.3%)	1		0.92
		A/G	236 (45.6%)	138 (47.9%)	1.10 (0.82-1.46)	0.54	94 (47.7%)	1.09 (0.78-1.51)	0.62	
	Recessive	A/A-A/G	414 (80.1%)	227 (78.8%)	1		156 (79.2%)	1		0.97
		G/G	103 (19.9%)	61 (21.2%)	1.08 (0.76-1.54)	0.67	41 (20.8%)	1.06 (0.70-1.59)	0.79	
	Dominant	A/A	178 (34.4%)	89 (30.9%)	1		62 (31.5%)	1		0.89
		A/G-G/G	339 (65.6%)	199 (69.1%)	1.17 (0.86-1.60)	0.31	135 (68.5%)	1.14 (0.80-1.62)	0.45	
	Codominant	A/A	178 (34.4%)	89 (30.9%)	1		62 (31.5%)	1		0.99
		A/G	236 (45.6%)	138 (47.9%)	1.17 (0.84-1.63)		94 (47.7%)	1.14 (0.79-1.66)		
		G/G	103 (19.9%)	61 (21.2%)	1.18 (0.79-1.78)	0.59	41 (20.8%)	1.14 (0.72-1.82)	0.75	
rs10941679	Overdominant	G/G-A/A	174 (45.3%)	126 (55%)	1		88 (54%)	1		0.75
		G/A	210 (54.7%)	103 (45%)	0.68 (0.49-0.94)	0.02	75 (46%)	0.71 (0.49-1.02)	0.063	
	Recessive	G/G-G/A	295 (76.8%)	176 (76.9%)	1		123 (75.5%)	1		0.84
		A/A	89 (23.2%)	53 (23.1%)	1.00 (0.68-1.47)	0.99	40 (24.5%)	1.08 (0.70-1.65)	0.73	
	Dominant	G/G	85 (22.1%)	73 (31.9%)	1		48 (29.4%)	1		0.61
		G/A-A/A	299 (77.9%)	156 (68.1%)	0.61 (0.42-0.88)	0.0081	115 (70.5%)	0.68 (0.45-1.03)	0.072	
	Codominant	G/G	85 (22.1%)	73 (31.9%)	1		48 (29.4%)	1		0.87
		G/A	210 (54.7%)	103 (45%)	0.57 (0.39-0.84)		75 (46%)	0.63 (0.41-0.98)		
		A/A	89 (23.2%)	53 (23.1%)	0.69 (0.44-1.10)	0.02	40 (24.5%)	0.80 (0.48-1.33)	0.12	
rs11878583	Overdominant	A/A-G/G	308 (59.3%)	169 (58.7%)	1		113 (56.8%)	1		0.58
		A/G	211 (40.7%)	119 (41.3%)	1.03 (0.77-1.38)	0.85	86 (43.2%)	1.11 (0.80-1.55)	0.53	
	Recessive	A/A-A/G	464 (89.4%)	262 (91%)	1		178 (89.5%)	1		0.68
		G/G	55 (10.6%)	26 (9%)	0.84 (0.51-1.37)	0.47	21 (10.6%)	1.00 (0.58-1.69)	0.99	
	Dominant	A/A	253 (48.8%)	143 (49.6%)	1		92 (46.2%)	1		0.46
		A/G-G/G	266 (51.2%)	145 (50.4%)	0.96 (0.72-1.29)	0.81	107 (53.8%)	1.11 (0.80-1.53)	0.55	
	Codominant	A/A	253 (48.8%)	143 (49.6%)	1		92 (46.2%)	1		0.72
		A/G	211 (40.7%)	119 (41.3%)	1.00 (0.74-1.35)		86 (43.2%)	1.12 (0.79-1.58)		
		G/G	55 (10.6%)	26 (9%)	0.84 (0.50-1.39)	0.77	21 (10.6%)	1.05 (0.60-1.83)	0.81	
rs7166081	Overdominant	G/G-A/A	295 (58.4%)	153 (53.3%)	1		107 (54.6%)	1		0.031
		A/G	210 (41.6%)	134 (46.7%)	1.23 (0.92-1.65)	0.16	89 (45.4%)	1.17 (0.84-1.63)	0.36	
	Recessive	G/G-A/G	378 (74.8%)	202 (70.4%)	1		155 (79.1%)	1		0.78
		A/A	127 (25.1%)	85 (29.6%)	1.25 (0.91-1.73)	0.17	41 (20.9%)	0.79 (0.53-1.17)	0.23	
	Dominant	G/G	168 (33.3%)	68 (23.7%)	1		66 (33.7%)	1		0.017
		A/G-A/A	337 (66.7%)	219 (76.3%)	1.61 (1.16-2.23)	0.0042	130 (66.3%)	0.98 (0.69-1.39)	0.92	
	Codominant	G/G	168 (33.3%)	68 (23.7%)	1		66 (33.7%)	1		0.022
		A/G	210 (41.6%)	134 (46.7%)	1.58 (1.10-2.25)		89 (45.4%)	1.08 (0.74-1.57)		
		A/A	127 (25.1%)	85 (29.6%)	1.65 (1.12-2.45)	0.016	41 (20.9%)	0.82 (0.52-1.29)	0.46	
rs16917302	Overdominant	A/A-C/C	338 (65.4%)	193 (67.2%)	1		128 (65.3%)	1		0.66
		C/A	179 (34.6%)	94 (32.8%)	0.92 (0.68-1.25)	0.59	68 (34.7%)	1.00 (0.71-1.42)	0.99	
	Recessive	A/A-C/A	496 (95.9%)	274 (95.5%)	1		186 (94.9%)	1		0.77
		C/C	21 (4.1%)	13 (4.5%)	1.12 (0.55-2.27)	0.75	10 (5.1%)	1.27 (0.59-2.75)	0.55	
	Dominant	A/A	317 (61.3%)	180 (62.7%)	1		118 (60.2%)	1		0.58
		C/A-C/C	200 (38.7%)	107 (37.3%)	0.94 (0.70-1.27)	0.69	78 (39.8%)	1.05 (0.75-1.47)	0.79	
	Codominant	A/A	317 (61.3%)	180 (62.7%)	1		118 (60.2%)	1		0.85
		C/A	179 (34.6%)	94 (32.8%)	0.92 (0.68-1.26)		68 (34.7%)	1.02 (0.72-1.45)		
		C/C	21 (4.1%)	13 (4.5%)	1.09 (0.53-2.23)	0.84	10 (5.1%)	1.28 (0.59-2.80)	0.83	

The *P* value is counted by the web-based tool SNPstats, the corresponding OR is counted after age adjustment

genotyping rate >80% and minor allele frequency (MAF) >0.01 was conducted. Statistical power of our study was shown in (Table 1). The MAF and OR was assumed at 0.03-0.30 and 1.1-2.0, respectively, and the power ranged from 0.0661 to 0.9999. In our study, 5 of 10 polymorphisms had a MAF greater than 0.15, only one SNP lower than 0.06. Therefore, the statistical power of this study was considered to be adequate.

Case patients and control subjects were comparable regarding age and geographical regions using a frequency–matched design (Table 2). The mean ages of patients and control subjects were 48.20 ± 9.850 and 46.99 ± 9.647 years, respectively. Additionally, the hormonal receptors of the case were distributed as follows: ER-positive in 288 cases, ER-negative in 199 cases, PR-positive in 273 cases and PR-negative in 214 cases (Table 2).

Of the 10 SNPs analyzed in the present case-control study, rs10941679 showed significant differences between the case and control groups (Table 3). The genotype distribution of the rs10941679 SNP in the breast cancer group was 22.3% GG, 54.7% GA and 23.1% AA, which was significantly different from that in the control group (30.09% GG, 45.4% GA and 23.7% AA; P = 0.012). Compared with GG and AA, the heterozygous genotype GA appeared to decrease breast cancer risk under the codominant (OR = 0.60, 95% CI = 0.43-0.84), dominant (OR = 0.64, 95% CI = 0.47–0.89) and overdominant models (OR = 0.69, 95% CI = 0.52–0.91). Among the remaining nine SNPs, four (rs311499 at 20q13.3, rs1045485 at CASP8, rs12964873 at CDH1, rs8170 at 19p13.1) showed no polymorphisms in the present population. No association was found among the remaining five SNPs.

Further stratification based on ER status identified

Table 5. SNPs that were Associated with PR-positive or PR-negative Tumours

				PR Positive			PR Negative			
SNP	Model	Genotype	Controls, N(%)	Cases, N (%)	OR(95% CI)	*P*	Cases, N (%)	OR(95% CI)	*P*	*P*-value
rs20775555	Overdominant	C/C-A/A	297 (57.2%)	147 (54.2%)	1		115 (54.5%)	1		0.97
		C/A	222 (42.8%)	124 (45.8%)	1.13 (0.84-1.52)	0.42	96 (45.5%)	1.12 (0.81-1.54)	0.5	
	Recessive	C/C-C/A	452 (87.1%)	251 (92.6%)	1		185 (87.7%)	1		0.069
		A/A	67 (12.9%)	20 (7.4%)	0.54 (0.32-0.91)	0.015	26 (12.3%)	0.95 (0.58-1.54)	0.83	
	Dominant	C/C	230 (44.3%)	127 (46.9%)	1		89 (42.2%)	1		0.26
		C/A-A/A	289 (55.7%)	144 (53.1%)	0.90 (0.67-1.21)	0.49	122 (57.8%)	1.09 (0.79-1.51)	0.6	
	Codominant	C/C	230 (44.3%)	127 (46.9%)	1		89 (42.2%)	1		0.16
		C/A	222 (42.8%)	124 (45.8%)	1.01 (0.74-1.38)		96 (45.5%)	1.12 (0.79-1.57)		
		A/A	67 (12.9%)	20 (7.4%)	0.54 (0.31-0.93)	0.052	26 (12.3%)	1.00 (0.60-1.68)	0.8	
rs12652447	Overdominant	A/A-G/G	282 (54.2%)	148 (54.4%)	1		104 (49.3%)	1		0.23
		A/G	238 (45.8%)	124 (45.6%)	0.99 (0.74-1.33)	0.96	107 (50.7%)	1.22 (0.88-1.68)	0.23	
	Recessive	A/A-A/G	417 (80.2%)	207 (76.1%)	1		175 (82.9%)	1		0.052
		G/G	103 (19.8%)	65 (23.9%)	1.27 (0.89-1.81)	0.18	36 (17.1%)	0.83 (0.55-1.27)	0.39	
	Dominant	A/A	179 (34.4%)	83 (30.5%)	1		68 (32.2%)	1		0.69
		A/G-G/G	341 (65.6%)	189 (69.5%)	1.20 (0.87-1.64)	0.27	143 (67.8%)	1.10 (0.79-1.55)	0.57	
	Codominant	A/A	179 (34.4%)	83 (30.5%)	1		68 (32.2%)	1		0.15
		A/G	238 (45.8%)	124 (45.6%)	1.12 (0.80-1.58)		107 (50.7%)	1.18 (0.83-1.70)		
		G/G	103 (19.8%)	65 (23.9%)	1.36 (0.91-2.04)	0.33	36 (17.1%)	0.92 (0.57-1.47)	0.45	
rs10941679	Overdominant	G/G-A/A	175 (45.3%)	123 (56.7%)	1		90 (51.7%)	1		0.31
		G/A	211 (54.7%)	94 (43.3%)	0.63 (0.45-0.89)	0.0074	84 (48.3%)	0.77 (0.54-1.11)	0.16	
	Recessive	G/G-G/A	297 (76.9%)	165 (76%)	1		134 (77%)	1		0.76
		A/A	89 (23.1%)	52 (24%)	1.05 (0.71-1.56)	0.8	40 (23%)	1.00 (0.65-1.52)	0.99	
	Dominant	G/G	86 (22.3%)	71 (32.7%)	1		50 (28.7%)	1		0.41
		G/A-A/A	300 (77.7%)	146 (67.3%)	0.59 (0.41-0.85)	0.0055	124 (71.3%)	0.71 (0.47-1.07)	0.1	
	Codominant	G/G	86 (22.3%)	71 (32.7%)	1		50 (28.7%)	1		0.58
		G/A	211 (54.7%)	94 (43.3%)	0.54 (0.36-0.80)		84 (48.3%)	0.68 (0.45-1.05)		
		A/A	89 (23.1%)	52 (24%)	0.71 (0.44-1.13)	0.0095	40 (23%)	0.77 (0.46-1.29)	0.23	
rs11878583	Overdominant	A/A-G/G	309 (59.2%)	163 (59.9%)	1		118 (55.4%)	1		0.36
		A/G	213 (40.8%)	109 (40.1%)	0.97 (0.72-1.31)	0.84	95 (44.6%)	1.17 (0.85-1.61)	0.34	
	Recessive	A/A-A/G	467 (89.5%)	249 (91.5%)	1		189 (88.7%)	1		0.3
		G/G	55 (10.5%)	23 (8.5%)	0.78 (0.47-1.31)	0.34	24 (11.3%)	1.08 (0.65-1.79)	0.77	
	Dominant	A/A	254 (48.7%)	140 (51.5%)	1		94 (44.1%)	1		0.13
		A/G-G/G	268 (51.3%)	132 (48.5%)	0.89 (0.67-1.20)	0.45	119 (55.9%)	1.20 (0.87-1.65)	0.26	
	Codominant	A/A	254 (48.7%)	140 (51.5%)	1		94 (44.1%)	1		0.27
		A/G	213 (40.8%)	109 (40.1%)	0.93 (0.68-1.27)		95 (44.6%)	1.21 (0.86-1.69)		
		G/G	55 (10.5%)	23 (8.5%)	0.76 (0.45-1.29)	0.57	24 (11.3%)	1.18 (0.69-2.01)	0.53	
rs7166081	Overdominant	G/G-A/A	297 (58.5%)	145 (53.7%)	1		115 (54.5%)	1		0.87
		A/G	211 (41.5%)	125 (46.3%)	1.21 (0.90-1.63)	0.2	96 (45.5%)	1.18 (0.85-1.62)	0.33	
	Recessive	G/G-A/G	380 (74.8%)	193 (71.5%)	1		162 (76.8%)	1		0.19
		A/A	128 (25.2%)	77 (28.5%)	1.18 (0.85-1.65)	0.32	49 (23.2%)	0.90 (0.62-1.31)	0.57	
	Dominant	G/G	169 (33.3%)	68 (25.2%)	1		66 (31.3%)	1		0.14
		A/G-A/A	339 (66.7%)	202 (74.8%)	1.48 (1.06-2.06)	0.019	145 (68.7%)	1.10 (0.78-1.55)	0.6	
	Codominant	G/G	169 (33.3%)	68 (25.2%)	1		66 (31.3%)	1		0.24
		A/G	211 (41.5%)	125 (46.3%)	1.47 (1.03-2.11)		96 (45.5%)	1.17 (0.80-1.69)		
		A/A	128 (25.2%)	77 (28.5%)	1.50 (1.00-2.23)	0.063	49 (23.2%)	0.98 (0.63-1.51)	0.62	
rs16917302	Overdominant	A/A-C/C	340 (65.4%)	179 (66.3%)	1		140 (66.3%)	1		0.98
		C/A	180 (34.6%)	91 (33.7%)	0.96 (0.70-1.31)	0.8	71 (33.6%)	0.96 (0.68-1.34)	0.8	
	Recessive	A/A-C/A	499 (96%)	256 (94.8%)	1		202 (95.7%)	1		0.64
		C/C	21 (4%)	14 (5.2%)	1.30 (0.65-2.60)	0.46	9 (4.3%)	1.06 (0.48-2.35)	0.89	
	Dominant	A/A	319 (61.4%)	165 (61.1%)	1		131 (62.1%)	1		0.82
		C/A-C/C	201 (38.6%)	105 (38.9%)	1.01 (0.75-1.37)	0.95	80 (37.9%)	0.97 (0.70-1.35)	0.85	
	Codominant	A/A	319 (61.4%)	165 (61.1%)	1		131 (62.1%)	1		0.89
		C/A	180 (34.6%)	91 (33.7%)	0.98 (0.71-1.34)		71 (33.6%)	0.96 (0.68-1.35)		
		C/C	21 (4%)	14 (5.2%)	1.29 (0.64-2.60)	0.76	9 (4.3%)	1.04 (0.47-2.34)	0.96	

The *P* value is counted by the web-based tool SNPstats, the corresponding OR is counted after age adjustment

two significant SNPs (Table 4). Compared to the CC-CA genotypes, the AA of rs2075555 in COL1A1 showed greater association with negative ER than with positive ER status (OR = 0.54; 95% CI, 0.29–0.99; *P* = 0.046). rs7166081 in SMAD3 also showed significant associations with ER status, but was associated with positive ER status in the recessive (OR = 1.59; 95% CI = 1.04–2.44; *P* = 0.031), codominant (OR = 2.01; 95% CI =1.22–3.33; P = 0.022) and dominant models (OR=1.64; 95% CI = 1.09–2.44; *P* = 0.017; Table 4). Although no significant association was found among these six SNPs and the PR status, marginal associations with PR status were observed among rs2075555 (OR = 0.57; 95% CI = 0.31-1.05; *P* = 0.069) and rs12652447 OR = 1.56; 95% CI =0.99–2.45; *P* = 0.052) in the recessive model (Table 5). We also evaluated the association between the SNPs and breast cancer risk in each subgroup which were stratified by the ER or PR status. Three breast cancer–associated SNPs, rs2075555, rs10941679 and rs7166081, in general, showed a stronger association with ER/PR-positive tumor than with ER/PR-negative (*P*<0.05) (Table 4, 5).

Discussion

Breast cancer is one of the most common cancers, and its incidence is ranked first among cancers in females (Jemal et al., 2011). Compared with western countries, China has a lower incidence of breast cancer (http://globocan.iarc.fr/factsheets/cancers/breast.asp). However, due to industry and lifestyle changes, the incidence of breast cancer in China is increasing, and breast cancer is being diagnosed in younger patients. In recent years, studies involving low-penetrance allelic variants have been conducted mainly using GWAS (Zheng et al., 2009;

Long et al., 2010; Haixin et al., 2013). These studies employed a large number of common genetic SNPs to identify associations with disease that rely on linkage disequilibrium (LD) patterns in the human genome. GWAS provides a powerful tool to identify novel markers for susceptibility and prognosis of disease. It has been used to identify new breast cancer susceptibility alleles by evaluating the association of genetic variants at various loci on different chromosomes (LD) in a large number of cases and controls by analyzing a panel of 1000 SNPs simultaneously. However, mixed populations may cause a potential false-positive type I error expansion due to multiple comparisons and other issues. High-quality validation studies are necessary to ensure that real associations with the disease are identified. Furthermore, as the majority of GWAS have been performed in European populations, the replication of GWAS-identified loci in specific Asian populations such as the Chinese or Japanese is also necessary (Easton et al., 2007; Hunter et al., 2007; Sueta et al., 2012). Therefore, we performed a verification study in a Chinese Han population using 10 susceptibility SNPs that had been reported previously in European populations.

Stacey et al. reported that the genetic susceptibility locus rs10941679 in chromosome 5p12 was associated with breast cancer risk (Stacey et al., 2008). Another study showed that this locus was associated with breast cancer risk only among females with more than 30 years of menstruation (odds ratio = 1.15, 95%CI: 1.05–1.26) (Zheng et al., 2009). In the present study, rs10941679 in 5p12/MRPS30 was associated with breast cancer risk under codominant (OR = 0.60; 95% CI = 0.43-0.84), dominant (OR= 0.64; 95%CI = 0.47-0.89), and overdominant models (OR = 0.69; 95%CI = 0.52–0.91). However, a fine-mapping analysis of 5p12 from the Shanghai Breast Cancer Study (Long et al., 2010) reported that rs10941679 showed a null association breast cancer risk (P=0.07), which was inconsistent with our finding. One reason for this difference may be that the samples in the Shanghai Breast Cancer Study just from shanghai. Meanwhile, the Chinese population samples of study carried by Chan et al. may from Southern Chinese descent (P=0.74) (Chan et al., 2012). Samples of present study are from Southern and Northern of China, respectively. The inconsistent result of these studies maybe mainly caused by the difference of geographic population.

Clinically, ER and PR status is an important biomarker to predict the response to endocrine treatment, which is the first and most efficacious target treatment for breast cancer with ER-positive and PR-positive histological markers (Mario et al., 2011). Breast cancer patients with ER- and/or PR-negative tumors experience higher mortality after their diagnosis compared with females with ER-positive and/or PR-positive disease (Anderson et al., 2001; Dunnwald et al., 2007). It is thought that variants in patients negative for these receptors could be related to the poor prognosis. Therefore, it is important to identify SNPs associated with breast cancer ER and PR positive/negative status. Our study revealed that rs2075555 in COL1A1 was associated with increased ER-negative breast cancer risk, and rs7166081 in SMAD3 appeared to be associated with ER-positive breast cancer. Using the generalized estimating equations (GEE) model of breast cancer, a previous report (Joanne et al., 2007) showed that rs2075555 in COL1A1 was associated with breast cancer ($p = 8.0 \times 10^{-8}$). However, another study (Daniele et al., 2011) reported that rs2075555 in COL1A1 did not show an association with breast cancer risk. Walker et al. provided evidence that SNPs rs7166081 in SMAD3 was associated with increased breast cancer risk for BRCA2 mutation carriers (OR = 1.25, 95% CI= 1.07–1.45, P = 0.004) (Walker et al., 2010). In our study, although rs2075555 in COL1A1 and rs7166081 in SMAD3 variants showed no associations with breast cancer, they were associated with the ER-negative/positive status in breast cancer, which may be useful for determination of the prognosis and molecular subtyping for endocrine treatment.

Fergus et al. found that rs16917302 in the ZNF365 locus was associated with both ER-positive and ER-negative status in BRCA2 mutation carriers (Fergus et al., 2012). Daniele et al. also show that SNP 5p12-rs10941679 was significantly associated with a greater risk of PR-positive than PR-negative breast cancer (P = 0.0028), which was inconsistent with our finding. One reason for this difference may be that the consortium included large well-established cohorts in the United States and Europe. Additionally, Asian females might have different lifestyles or environmental exposures that affect their risk of breast cancer. Genetic interactions with other SNPs that differ in frequency between populations could also manifest as a heterogeneity effect.

In conclusion, this was a verification study in a Chinese Han population using 10 susceptibility SNPs reported previously in European populations. However, only one SNP (rs10941679 in 5p12) was confirmed, indicating differences between European and Chinese populations in terms of breast cancer susceptibility loci. Therefore, confirmation studies are necessary before these loci in can be applied in Chinese populations.

Acknowledgements

This research is supported by the Doctoral Program of Higher Education Foundation (Grant No. 20104433120016) and the National Natural Science Foundation of China (Grant No. 81302327).

References

Anderson WF, Chu KC, Chatterjee N, et al (2001). Tumor variants by hormone receptor expression in white patients with node-negative breast cancer from the surveillance, epidemiology, and end results database. *J Clin Oncol*, **19**, 18-27.

Campa D, Kaaks R, Le Marchand L, et al (2011). Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *J Natl Cancer Inst*, **103**, 1252-63.

Chan M, Ji SM, Liaw CS, et al (2012). Association of common genetic variants with breast cancer risk and clinic opathological characteristics in a Chinese population. *Breast Cancer Res Treat*, **136**, 209-20.

Couch FJ, Gaudet MM, Antoniou AC, et al (2012). Common

variants at the 19p13.1 and ZNF365 loci are associated with ER subtypes of breast cancer and ovarian cancer risk in BRCA1 and BRCA2 mutation carriers. *Cancer Epidemiol Biomarkers Prev*, **21**, 645-57.

Dunnwald LK, Rossing MA, Li CI (2007). Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res,* **9**, R6.

Easton DF, Pooley KA, Dunning AM, et al (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, 1087-93.

Li H, Beeghly-Fadiel A, Wen W, et al (2013). Gene-environment interactions for breast cancer risk among Chinese women: a report from the Shanghai Breast Cancer Genetics Study. *Am J Epidemiol*, **177**, 161-70.

Hunter DJ, Kraft P, Jacobs KB, et al (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*, **39**, 870-4.

Jemal A, Bray F, Center MM, et al (2011). Global cancer statistics. *CA Cancer J Clin*, **61**, 69-90.

Jemal A, Siegel R, Ward E, et al (2006). Cancer statistics. *CA Cancer J Clin*, **56**, 106-30.

Murabito JM, Rosenberg CL, Finger D, et al (2007). A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Medical Genetics*, **8**, S6

Long J, Shu XO, Cai Q, et al (2010). Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol Biomarkers Prev*, 19, 2357-65.

Mario, G., Rachel, SC., Kent, O., et al. (2011). Biological mechanisms and clinical implications of endocrine resistance in breast cancer. *The Breast*, **20S3**, S42-9.

McCracken M, Olsen M, Chen MS Jr, et al (2007). Cancer incidence, mortality, and associated risk factors among Asian Americans of Chinese, Filipino, Vietnamese, Korean, and Japanese ethnicities. *CA Cancer J Clin*, **57**, 190-205.

Pharoah PD, Antoniou AC, Easton DF, et al (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med*, **358**, 2796-803.

Raskin L, Pinchev M, Arad C, et al (2008). FGFR2 is a breast cancer susceptibility gene in Jewish and Arab Israeli populations. *Cancer Epidemiol Biomarkers Prev*, **17**, 1060-5.

Stacey SN, Manolescu A, Sulem P, et al (2008). Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*, **40**, 703-6.

Sueta A, Ito H, Kawase T, Hirose K, et al (2012). A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. *Breast Cancer Res Treat*, **132**, 711-21.

Walker LC, Fredericksen ZS, Wang X, et al (2010). Evidence for SMAD3 as a modifier of breast cancer risk in BRCA2 mutation carriers. *Breast Cancer Res*, **12**, R102.

Walsh T, King MC (2007). Ten genes for inherited breast cancer. *Cancer Cell*, **11**, 103-5.

Yu KD, Di GH, Wu J, et al (2007). Development and trends of surgical modalities for breast cancer in China: a review of 16- year data. *Ann Surg Oncol*, **14**, 2502-9.

Zheng W, Long J, Gao YT, et al (2009). Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet*, **41**, 324-8.